

AMENDMENTS TO THE CLAIMS

Claim 1 (Cancelled)

Claim 2 (Cancelled)

Claim 3 (Original) Method for solving a community generation problem, comprising the steps of:

converting documents to digital form and tagging said digitized documents;
parsing said digitized and tagged documents to extract the transaction history vector for each individual;

creating timelines of said transaction vectors so as to form a timeline map;

determining the relevancy of said vectors;

projecting said vectors along a time dimension so as to form a histogram;

translating said vectors into groups of activities by histogram clustering;

determining the local correlation between any pair of clusters in the timeline of two individuals;

computing the global correlations between pairs of individuals;

converting data to a graph as a function of all individuals extracted from said documents and the correlation values between said individuals;

generating models based on a search of all subgraphs with correlation values above a threshold; and

outputting a group model.

Claim 4 (Original) Method of claim 3, wherein said step of parsing further comprises the step of applying the “one way nearest neighbor” principle.

Claim 5 (Original) Method of claim 4, wherein said “one way nearest neighbor” principle further comprises the following steps as applied to a money laundering problem:

for every person's name encountered, the first immediate time instance is the first time instance for a series of financial activities; the second immediate time instance is the second time instance for another series of financial activities, etc.;

for every time instance encountered, all the subsequent financial activities are considered as the series of financial activities between this time instance and the next time instance;

financial activities are identified in terms of money amount; money amount is neutral in terms of deposit or withdrawal;

each person's time sequence of financial activities is updated if new financial activities of this person are encountered in other places of the same document or in other documents; and

the financial activities of each time instance of a person is updated if new financial activities of this time instance of the same person are encountered in other places of the same document or in other documents.

Claim 6 (Original) Method of claim 3, wherein said step of determining the relevancy of said vectors further comprises a step of focusing on "clusters" of vectors in said timeline map and ignoring scattered (i.e., non-clustered) vectors in said timeline map.

Claim 7 (Original) Method of claim 3, wherein said step of translating said vectors into groups of activities further comprises solving a standard histogram clustering problem; and

simplifying said standard clustering problem by virtue of all individuals sharing the same said timeline.

Claim 8 (Original) Method of claim 3, wherein said step of computing correlations between pairs of individuals further comprises computing the global correlation of all local correlations between pairs of individuals.

Claim 9 (Original) Method of claim 8, further comprising the step of computing local correlations by computing the correlation between two clusters corresponding to a pair of individuals on said histograms.

Claim 10 (Cancelled)

Claim 11 (Cancelled)

Claim 12 (Original) Method of claim 8 wherein said step of computing the global correlation of all local correlations between pairs of individuals further comprises computing the dot product between two vectors as follows:

$$C(x, y) = Cy(x) \cdot Cx(y) = \sum_{i=1}^K C(x_i, y) C(y_i, x)$$

where the vectors $Cy(x)$ and $Cx(y)$ are defined as

$$Cy(x) = \langle C(x_i, y), i = 1, \dots, K \rangle$$

$$Cx(y) = \langle C(y_i, x), i = 1, \dots, K \rangle$$

where $C(x_i, y) = \max_{j=1}^K \{g(x_i, y_j) S(i, j)\}$

$$S(i, j) = e^{-\frac{(c_i - c_j)^2}{2\sigma_i^2}}$$

and where

$$\{g(x_i, y_j) S(i, j), j = 1, \dots, K\}$$

Claim 13 (Original) Method of claim 3, wherein said step of converting data to a graph further comprises obtaining a complete graph $G(V, E)$, where V is the set of all the individuals extracted from the given collection of the documents, and E is the set of all the correlation values between individuals such that for any correlation $C(x, y)$, there is a corresponding edge in G with the weight C between the two nodes x and y .

Claim 14 (Original) Method of claim 3, wherein said step of generating models further comprises the step of identifying links as a graph segmentation based on a minimum correlation threshold value.

Claim 15 (Original) Method of claim 14, wherein said minimum threshold value is selected based upon a user's expertise.